Hierarchical Estimation as Basis for Hierarchical Forecasting

Strijbosch, L.W.G.; Heuts, R.M.J.; Moors, J.J.A.

Publication date:
2006

CentER

Discussion Paper

No. 2006–86

HIERARCHICAL ESTIMATION AS BASIS FOR HIERARCHICAL FORECASTING

By L.W.G. Strijbosch, R.M.J. Heuts, J.J.A. Moors

September 2006

ISSN 0924-7815

TILBURG UNIVERSITY

Hierarchical estimation as basis for hierarchical forecasting

L.W.G. Strijbosch, R.M.J. Heuts & J.J.A. Moors

Tilburg University,

CentER

Department of Econometrics & Operations Research

Warandelaan 2, P.O. Box 90153,

5000LE Tilburg, The Netherlands

Abstract

In inventory management, hierarchical forecasting (HF) is a hot issue: families of items are formed for which total demand is forecasted; total forecast then is broken up to produce forecasts for the individual items. Since HF is a complicated procedure, analytical results are hard to obtain; consequently, most literature is based on simulations and case studies.

This paper succeeds in following a more theoretical approach by simplifying the problem: we consider estimation instead of forecasting. So, from a random sample we estimate both total demand and the fraction of this total that individual items take; multiplying these two quantities gives a new estimate of individual demand. Then our research question is: can aggregation of items, followed by fractioning, lead to more accurate estimates of individual demand?

We consider two simple situations that can be analysed fully theoretically. Thirdly, a more practical situation is investigated by means of simulation.

JEL-code: C53

Keywords: hierarchical forecasting; aggregation, top-down approach

1 Introduction

We will consider n individual items i, for which demand in a fixed period is a random variable X_i $(i = 1, 2, \ldots, n)$. We will assume that the general form of their joint distribution D is known; however, D contains unknown parameters. This holds as well for the marginal distributions D_i so that the individual means $\mu_i = E(X_i)$ and variances $\sigma_i^2 = V(X_i)$ $(i = 1, 2, \ldots, n)$ are unknown as well. We want to estimate the means from a random sample of size m.

The straightforward estimator for each μ_i of course is the sample mean $\bar{X}_i-$ with variance $\sigma_i^2/m-$ which ignores all interrelations between the items. Here, another estimator will be introduced that tries to take advantage from these interrelationships.

For the n items, total demand Y and fractional demand Z_i can be written as

$$Y = \sum\nolimits_{i=1}^{n} X_i, \quad Z_i = X_i/Y$$

We separately estimate the expectations of these two random variables by the sample means $\bar{Y}$ and $\bar{Z}_i$, respectively. Our new top-down hierarchical estimators S_i for μ_i then are defined as

$$S_i = \bar{Y} * \bar{Z}_i, \; i = 1, 2, \ldots, n$$

Without loss of generality we will concentrate on the first item and write $Z = Z_1$ and $S = S_1$ for simplicity; hence, in the sequel, attention will be concentrated on the estimator

$$S = \bar{Y} * \bar{Z} \tag{1}$$

Note that the distributions of Y and Z, and consequently of $\bar{Y}, \bar{Z}$ and S, are completely determined by D. But of course, in general no analytical expressions will exist for these distributions, so that it is difficult to compare the performance of S and $\bar{X}_1$.

Therefore, we start with two special, relatively simple cases, for which expectation and variance of S can be found theoretically; then, the accuracy of S and $\bar{X}_1$ can be compared. Our third case is more practice-oriented and hence more complicated; therefore, comparisons in performance will be based on simulations.

The remainder of the paper is organised as follows. Section 2 presents an overview of recent literature on Hierarchical Forecasting - in particular the top-down approach – that triggered this paper.

In Section 3 we give the answer to our research question in the special case that all items i have independent demands, all having a gamma distribution with the same scale parameter. We will show that now our hierarchical estimator S is never less accurate than $\bar{X}_1$. In words: estimation of mean item demand can be improved with the help of other items with *fully independent demand*! To phrase it more popular: using demand data for bicycle parts in Marciac, France, improves estimation of bakery products in Tilburg, The Netherlands.

Section 4 treats the case of just two items with demand equalling either 1 or 2. Introducing three parameters makes this simple situation nevertheless interesting, because demand for the two items now may be (negatively or positively) dependent. Now, S performs better for certain parameter combinations only.

Since gamma distributions are widely accepted as a very fruitful demand model, we consider this situation again in Section 5. Here, individual demand X_i may have any gamma distribution, while dependence is allowed. Our simulation results show that for $n = 2$, the hierarchical estimator is better if the scale parameter of X_1 is the larger of the two.

The final Section 6 summarizes and discusses our results.

2 Literature review on hierarchical forecasting

For production-inventory planning purposes many firms have to rely on short-term demand forecasts. This means that not just forecasting an average demand is relevant, but also forecasting its accuracy.

As these forecasts often have to been generated daily or weekly for thousands of items, this means that management is faced with a burdensome task. Given that the product life cycles become shorter, demand patterns more irregular and often too few data are available, it is nearly unfeasible to build models for individual products. Therefore, family-based forecast systems are developed, based upon a user-defined strategy for aggregating items into families, attempting to improve forecast performance and at the same time reducing the overall forecast burden.

In the so-called Hierarchical Forecasting (HF) literature, which aims at providing forecasts of aggregate and individual demands through either a 'bottom-up' or a 'top-down' forecast process, one tries to analyse what's the best approach in what situation.

In the top-down forecast procedure, first a direct forecast of the aggregate is determined and then this aggregate forecast is used to determine individual demand forecasts. For example, one could multiply the aggregate demand forecast by the forecast of the ratio of the respective individual item demand to aggregate demand - called *proration* in the literature (see e.g. Fliedner (1999) for recent information on HF). This approach was already successfully applied by Steece and Wood (1979), who derived also an approximation for the standard errors of these top-down forecasts - important for inventory management purposes.

The applicability of these methods seems dependent on four factors: (1) the right criterion for constructing a family of items based on similar demand items, (2) the accuracy of the total demand aggregate forecast, (3) the accuracy of the fraction or proportion used in conjunction with the size of the individual product's share of the total demand sales, and (4) the forecast techniques used for the aggregate and the fraction. For example, Gross and Sohl (1990) demonstrated, in a large empirical study, that a simple sample average for the proration factor worked relative well compared to several other methods. But they also stated: 'The larger a product's share, the greater the accuracy required for the estimated proportions'.

In the bottom-up approach, individual forecasts for each item demand are added to produce a forecast of aggregate demand.

Fliedner and Lawrence (1995) suggested that the method to identify group candidates is not really important. They observed that sophisticated statistical clustering techniques used to identify family members did not lead to better forecast performance than when families were randomly determined. However, Zotteri et al. (2005) show that what is driving the HF forecast accuracy is the choice of the appropriate level of aggregation (e.g. store level, distribution centre level or chain level).

The literature on HF can be divided into two main streams concerning the two approaches. A group of authors argue that the top-down approach is superior because of its lower cost and greater accuracy, at least for stationary demand. Among them are Plossl (1973) and Ilmakunnas (1990). Lapide (1998) discusses situations where top-down is preferable and situations where bottom-up makes sense. Recently, Dekker et al. (2004) demonstrated the advantage of aggregation to improve Winters' seasonal demand forecast procedure. However, their simulations are based on aggregating similar items, each with the same gamma demand distribution and the same seasonal pattern and then it is logical that one gets better forecasts. Zellner and Tobias (2000) in a case study provide evidence that improved forecasting results can be obtained by disaggregation, given that an aggregate variable appears in the disaggregated relations.

Contrary to that finding Dangerfield and Morris (1992) found, using a subset of the M-competition database, that bottom-up forecasting of family items produces more accurate forecasts. Also Shlifer and Wolff (1979) state that there are reasonable conditions under which bottom-up is to be preferred over direct forecasting an aggregate variable. Dunn et al. (1976) concluded from a case study that aggregated forecasts based on forecasts from geographically smaller wire center areas were better when compared with those developed directly from aggregated data on level area.

Another group of authors, such as Aigner and Goldfeld (1974), Wei and Bovas (1981), Schwarzkopf et al. (1988), outline the relative advantages of each strategy. To their opinion, favouring one strategy over the other very much depends on the modelling assumptions, such as correlation structure on item level, parameter estimation procedures, and sampling bias. Authors like Miller et al. (1976), Barnea and Lakonishok (1980), and Fliedner (1999) demonstrate that the forecast performance between top-down vs. bottom-up strategies is dependent on: (1) the magnitude of the correlation between the individual variables, (2) the magnitude of the correlation between forecast errors of individual variables, and (3) the quality of the forecasting technique(s). Fliedner (1999, 2001) further observed that high positive and high negative correlation between the items afforded improved forecast performance at the aggregate level. In his opinion items possessing complimentary or contra-cyclical demand patterns may be considered for inclusion within a family.

3 Independent gamma distributions

In this first estimation problem, item demand X_i has a gamma distribution, with the same scale parameter λ for all i. Further, it will be assumed that all X_i are statistically independent. So, if $\Gamma(p,\lambda)$ denotes the gamma distribution with shape parameter p and scale parameter λ, the joint distribution D is fully described by

$$\begin{cases} X_i \sim \Gamma(p_i,\lambda), \; i=1,2,\ldots,n \\ X_i, X_j \text{ independent for all } i \neq j \end{cases} \tag{2}$$

We will take $\lambda = 1$ without loss of generality. A slight extension of Theorem 1.2.3 in Bickel and Doksum (1977) gives that the joint distribution of total demand Y and fractional demand Z is given by

$$\begin{cases} Y \sim \Gamma(p_1+q,1) \\ Z \sim B(p_{1,}q) \\ Y,\, Z \text{ independent} \end{cases}$$

where B indicates the beta distribution and $q = \sum_{i=2}^{n} p_i$. Some direct consequences are

$$\begin{cases} E(Y) = V(Y) = p_1 + q \\ E(Z) = p_1/(p_1+q), V(Z) = p_1 q / \left\{(p_1+q)^2 (p_1+q+1)\right\} \end{cases} \tag{3}$$

If all X_i are observed during m random periods, the sample mean $\bar{X}_i$ has expectation p_i and variance $V(X_i) = p_i/m$. Similar results hold for Y and $\bar{Z}$.

Because of the independence of Y and Z, $S = \bar{Y} * \bar{Z}$ has the moments

$$E(S) = E(\bar{Y})E(\bar{Z})$$

$$V(S) = V(\bar{Y})V(\bar{Z}) + E^2(\bar{Y})V(\bar{Z}) + E^2(\bar{Z})V(\bar{Y})$$

Plugging in (3) shows that the hierarchical estimator S is unbiased for $E(X_1) = p_1$ with

$$V(S) = p_1 \left[p_1 + (p_1+q)^2 + q/m\right] / \left[m\,(p_1+q)\,(p_1+q+1)\right] \tag{4}$$

Define the gain G in accuracy of S with respect to $\bar{X}_1$ by

$$G = V(\bar{X}_1) - MSE(S) \tag{5}$$

where MSE denotes the Mean Squared Error. Then here we obtain

$$G = p_1/m - V(S) = \frac{m-1}{m^2} \frac{p_1 q}{(p_1+q)(p_1+q+1)} \tag{6}$$

so that S is more accurate than $\bar{X}_1$ for all $m > 1$.

So, we have the paradoxical situation that estimation of item demand is improved by using observed demand of other, *fully unrelated,* items. We have no explanation for this phenomenon. It might be thought that it is caused by the notoriously difficult problem of estimating gamma distribution parameters (see e.g. Bowman and Shenton, 1988). But here only the gamma mean is estimated, for which no better estimator has been proposed than the sample mean - our $\bar{X}_1$; see for a recent discussion Balakrishnan and Wang (2000).

4 Two dichotomous variables

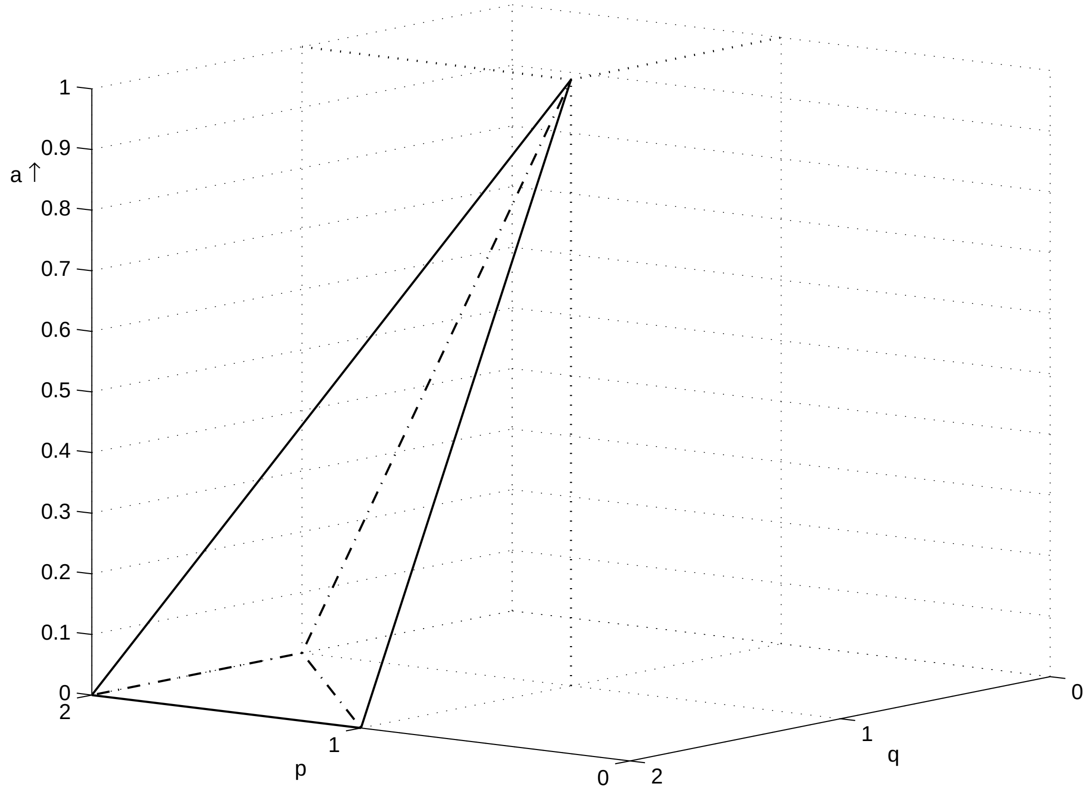
In this section we consider the simple case of just two items with demand X_1 and X_2, taking 1 and 2 as possible values; they are observed during two random periods, so $n = m = 2$. Their joint distribution $D(a, p, q)$, depending on three parameters, is shown in Table 1; the entries are the joint probabilities $p(x_1, x_2)$.

Table 1. Joint distribution $D(a, p, q)$ of (X_1, X_2).

x_2 / x_1	1	2	Total
1	a	$2-a-p$	$2-p$
2	$2-a-q$	$a+p+q-3$	$p-1$
Total	$2-q$	$q-1$	1

Of course, all entries should be non-negative; the resulting parameter space is pictured in Figure 1. For any given a, the values of (p, q) are restricted to the triangle with the corner points $(1, 2-a)$, $(2-a, 2-a)$ and $(1, 2-a)$.

Figure 1. Parameter space of $D(a, p, q)$.



It follows immediately that $\mu_1 = p$; hence, p is the parameter of interest. Other moments are:

$$\sigma_1^2 = V(X_1) = (2-p)(p-1)$$
$$\sigma_{12} = Cov(X_1, X_2) = a - (2-p)(2-q)$$

so that the correlation coefficient can take all values in $[-1, +1]$.

The consequent joint distribution of total demand Y and fractional demand $Z = X_1/Y$ is presented in Table 2.

Table 2. Joint distribution of (Z, Y).

z \ y	2	3	4	Total
$\frac{1}{3}$	0	$2-a-p$	0	$2-a-p$
$\frac{1}{2}$	a	0	$a+p+q-3$	$2a+p+q-3$
$\frac{2}{3}$	0	$2-a-q$	0	$2-a-q$
Total	a	$4-2a-p-q$	$a+p+q-3$	1

From the $m = 2$ observations of (X_1, X_2) the averages $(\bar{Z}, \bar{Y})$ can be calculated; for simplicity, Table 3 shows the distribution of $(V, W) = (12\bar{Z} - 6, 2\bar{Y} - 6)$; only positive values of the joint probability function $p(v, w)$ are shown.

Table 3. Joint probabilities $p(v, w)$ of (V, W).

(v, w)	$p(v, w)$
$(-2, 0)$	$(2-a-p)^2$
$(-1, -1)$	$2a(2-a-p)$
$(-1, 1)$	$2(2-a-p)(a+p+q-3)$
$(0, -2)$	a^2
$(0, 0)$	$2(2-a-p)(2-a-q)+2a(a+p+q-3)$
$(0, 2)$	$(a+p+q-3)^2$
$(1, -1)$	$2a(2-a-q)$
$(1, 1)$	$2(2-a-q)(a+p+q-3)$
$(2, 0)$	$(2-a-q)^2$

Writing

$$v = p - q, s = p + q - 3$$

for brevity, the bias $B(S)$ of our hierarchical estimator S becomes:

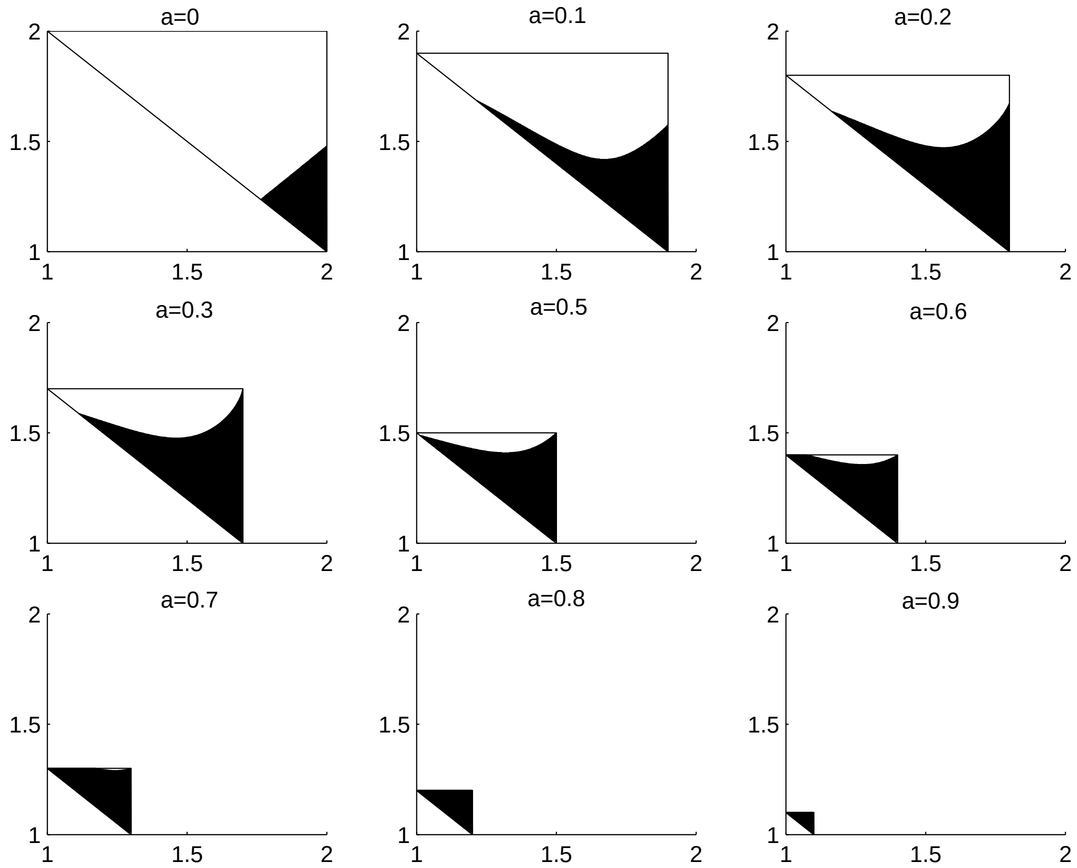
$$B(S) = E(S) - p = vs/12 \tag{7}$$

Straightforward but somewhat elaborate algebra gives:

$$288G = 2a\left[14s - 12v + 2a - 1\right] + s\left[24v\left(s + v\right) + 13s - 12v - 13\right] \tag{8}$$

(Some intermediate results are summarized in Appendix 1.) Our new estimator is preferable whenever the gain G is positive; Figure 2 shows when this occurs, depending on a.

Figure 2. Parameter values for which S is preferable to X_1.



It is clear that in large parts of the parameter space, the new estimator outperforms the standard estimator $\bar{X}_1$; the relative size of this part increases with a, hence with the covariance.

5 Two arbitrary gamma distributed variables

Like in Section 3, we consider gamma distributed individual demand X_i, allowing now, however, both varying scale parameters, and dependence. On the other hand, we restrict ourselves to the case of $n = 2$ items. Starting from $X_1 \sim \Gamma(p_1, \lambda_1)$ we will construct another (dependent) gamma variable $X_2 \sim \Gamma(p_2, \lambda_2)$.

Indicate the median of $\Gamma(p_i, \lambda_i)$ by M_i. Let $\Gamma^-(p_i, \lambda_i)$ and $E^-(X_i)$ denote the conditional distribution and expectation of X_i given $X_i < M_i$, and similarly $\Gamma^+(p_i, \lambda_i)$ and $E^+(X_i)$ for $X_i > M_i$. Now, for any random observation $X_1 \sim \Gamma(p_1, \lambda_1)$, an observation of X_2 is constructed as follows:

$$\text{if } X_1 > M_1, \text{ then } X_2 \begin{cases} \sim \Gamma^-(p_2, \lambda_2) \text{ with probability } \alpha \\ \sim \Gamma^+(p_2, \lambda_2) \text{ with probability } 1-\alpha, \end{cases}$$
$$\text{if } X_1 < M_1, \text{ then } X_2 \begin{cases} \sim \Gamma^-(p_2, \lambda_2) \text{ with probability } 1-\alpha \\ \sim \Gamma^+(p_2, \lambda_2) \text{ with probability } \alpha. \end{cases}$$

It will be clear that indeed $X_2 \sim \Gamma(p_2, \lambda_2)$. Figure 3 shows the probabilities that (X_1, X_2) takes values in one of the four subspaces of the first quadrant, plus the conditional expectations of X_1X_2.

Figure 3. Conditional expectations of X_1X_2.

$x_2 \uparrow$			
$x_2 > M_2$	$\alpha/2$ $E^-(X_1)\,E^+(X_2)$	$(1-\alpha)/2$ $E^+(X_1)\,E^+(X_2)$	
$x_2 < M_2$	$(1-\alpha)/2$ $E^-(X_1)\,E^-(X_2)$	$\alpha/2$ $E^+(X_1)\,E^-(X_2)$	
0	$x_1 < M_1$	$x_1 > M_1$	$x_1 \longrightarrow$

It is clear that X_1, X_2 are positively (negatively) correlated for $\alpha < 0.5$ (> 0.5), while $\alpha = 0.5$ quarantees independence. Indeed

$$Cov(X_1, X_2) = (1-2\alpha)\left[E(X_1) - E^-(X_1)\right]\left[E(X_2) - E^-(X_2)\right]$$

Note that replacing α by $1-\alpha$ changes the sign of $Cov(X_1, X_2)$. Table 4 shows some simulated values, based on $50,000$ pairs of observations of (X_1, X_2).

Table 4. Simulated values of $Cov(X_1, X_2)$.

	p_2	0.5			1			5		
α	p_1	0.5	1	5	0.5	1	5	0.5	1	5
0.1		0.294	0.336	0.375	0.339	0.386	0.428	0.379	0.430	0.481
0.3		0.149	0.168	0.188	0.169	0.193	0.214	0.191	0.215	0.241
0.5		0.002	0.004	−0.001	0.001	0.001	−0.001	0.002	0.003	−0.001
0.7		−0.144	−0.170	−0.186	−0.168	−0.193	−0.213	−0.188	−0.216	−0.237
0.9		−0.296	−0.338	−0.376	−0.338	−0.386	−0.431	−0.378	−0.433	−0.483

All covariances obtained in this way range from -0.483 to 0.481. More extreme values could be obtained by considering quartiles of X_1 and X_2 instead of just the medians. We give a simple example. Divide $\mathbb{R}^+$ in four parts I_i $(i = 1, ..., 4)$ with the three quartiles of X_1 as division points; another subdivision J_i $(i = 1, ..., 4)$ of $\mathbb{R}^+$ is obtained by means of the quartiles of X_2. Now, a random observation X_2, depending on X_1, is constructed as follows:

if $X_1 \in I_i$, X_2 is drawn at random from J_{5-i} for $i = 1, ..., 4$.

For the dependent pair (X_1, X_2) constructed above we simulated the gain of the hierarchical estimator S. We took $p_1 = \lambda_1 = 1$, so X_1 has the (standard) exponential distribution from now on. Both p_2 and λ_2 were taken from $\{0.5, 1, 5\}$ and we used $m \in \{2, 5, 10\}$ observations of (X_1, X_2). The gain G, defined in (5), was calculated from $100,000$ simulations. Table 5 shows the results.

Table 5. Gain $(\times 10^4)$ of hierarchical estimator S w.r.t. $\overline{X}_1$.

	λ_2	0.5			1			5			
α	p_2	0.5	1	5	0.5	1	5	0.5	1	5	m
0.1		128	330	681	40	236	564	-4293	-2694	-90	
0.3		242	439	681	208	404	397	-4261	-3560	-684	
0.5		257	451	609	275	410	260	-4485	-4640	-1441	2
0.7		382	562	663	517	502	231	-4284	-5338	-1942	
0.9		501	640	755	596	674	232	-4079	-5618	-2354	
0.1		104	202	342	26	144	287	-3892	-2044	-29	
0.3		166	264	365	130	239	254	-3949	-2880	-479	
0.5		173	267	369	201	246	171	-4218	-3938	-1005	5
0.7		234	323	409	294	324	138	-4167	-4840	-1517	
0.9		242	345	461	356	389	89	-4145	-5719	-2071	
0.1		69	108	92	-2	78	105	-3285	-1561	-12	
0.3		92	122	147	71	128	129	-3419	-2328	-281	
0.5		87	125	180	114	144	102	-3680	-3356	-691	10
0.7		97	140	218	161	191	60	-3787	-4406	-1161	
0.9		71	136	257	179	218	-7	-3816	-5453	-1723	

The obvious conclusion is that S outperforms $\overline{X}_1$ whenever $\lambda_2 \leq \lambda_1 (= 1)$; this holds irrespective of the values of the threesome (p_2, α, m). Again, we have no explanation of this phenomenon.

6 Conclusion

In estimating mean demand of individual items, other items were included in the estimation procedure. In particular, we considered the (top-down) hierarchical estimator S, obtained by multiplying the separate sample means of total and fractional demand. We showed that this hierarchical estimator may outperform the sample mean of the item demand. This surprising result was proved in a specific case of gamma distributed demand; simulation showed that it holds in more general situations as well, both for negative and positive correlations between individual item demands. A partial explanation is that total (fractional) demand has smaller variance in case of negative (positive) dependence.

That the hierarchical estimator often is preferable, apart from the gamma distribution, was shown in Section 4 where a totally different joint distribution was considered.

Appendix

From Table 3, the following moments of (V, W) are derived easily.

$$E\left(V^2\right) = 2\left[1 - 2a - s + v^2\right]$$

$$E(W^2) = 2\left[2a + s(s+1)\right]$$

$$E(VW) = 2sv$$

$$E(V^2W) = 2s\left[1 - s - 2a\right]$$

$$E(VW^2) = 2v\left[2a + s\right]$$

$$E(V^2W^2) = 2\left[1 - s - 2a\right]\left[2a + s\right].$$

Plugging in these results in

$$MSE(S) = E(S^2) - 2pE(S) + p^2$$

gives (6).

References

Aigner DJ, Goldfeld SM. Estimation and prediction from aggregate data when aggregates are measured more accurately than their components. Econometrica 1974; 42, No.1: 113-134.

Balakrishnan N, Wang J. Simple efficient estimation for the three parameter gamma distribution. Journal of Statistical Planning and Inference 2000; 85: 115-126.

Barnea A, Lakonishok J. An analysis of the usefulness of disaggregated accounting data for forecasts of corporate performance. Decision Sciences 1980; 11: 17-26.

Bickel PJ, Doksum KA. Mathematical Statistics: Basic Ideas and selected Topics. Holden-Day, San Francisco, 1977.

Bowman KO, Shenton LR. Properties of Estimators for the Gamma Distribution. Marcel Dekker, New York, 1988.

Dangerfield BJ, Morris JS. Top-down or bottom-up: aggregate versus disaggregate extrapolations. International Journal of Forecasting 1992; 8, Issue 2: 233-241.

Dekker M, Donselaar K, Ouweland P. How to use aggregation and combined forecasting to improve seasonal demand forecasts. International Journal of Production Economics 2004; 90, Issue 2: 151-167.

Dunn DM, Williams WH, DeChaine TL. Aggregate versus subaggregate models in local area forecasting. Journal of the American Statistical Association 1976; 71; No. 353: 68-71.

Fliedner G. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. Computers and Operations Research 1999; 26: 1133-1149.

Fliedner G. Hierarchical forecasting: issues and use guidelines. Management Systems and Data Systems 2001; 101; Issue 1: 1-6.

Fliedner EB, Lawrence B. Forecasting system parent group formation: An empirical application of cluster analysis. Journal of Operations Management 1995; 12: 119-130.

Gross CW, Sohl JE. Disaggregation methods to expedite product line forecasting. Journal of Forecasting 1990; 9: 233-254.

Ilmakunnas P. Aggregation vs. disaggregation in forecasting construction activities, in TS Barker and H Pesaran, Disaggregation in Econometric Modelling. Routledge, London: 1990, 73-86.

Lapide L. New developments in business forecasting. Journal of Business Forecasting 1998; Summer: 28-29.

Miller JG, Berry W, Lai C-YF. A comparison of alternative forecasting strategies for multi-stage production inventory systems. Decision Sciences 1976; 7: 714-724.

Plossl GW. Getting the most from forecasts. Production and Inventory Management 1973; 14:1-16.

Schwarzkopf AB, Tersine J, Morris JS. Top-down versus bottom-up forecasting strategies. International Journal of Production Research 1988; 26, No. 11: 1833-1843.

Shlifer E, Wolff RW. Aggregation and proration in forecasting. Management Science 1979; 25, No.6: 594-603.

Steece BM, Wood SD. An ARIMA-based methodology for forecasting in a multi-item environment, in S Makridakis and SC Wheelwright, Forecasting, vol. 12, Studies in the Management Sciences, North Holland, Amsterdam, 1979.

Wei W, Bovas A. Forecasting contemporal time series aggregates. Communications in Statistics A 1981; 10: 1335-1344.

Zellner A, Tobias J. A note on aggregation, disaggregation and forecasting performance. Journal of Forecasting 2000, 19, Issue 5: 457-465.

Zotteri G, Kalchschmidt M, Caniato F. The impact of aggregation level on forecasting performance. International Journal of Production Economics 2005; 93-94: 479-491.